Making Automation Explicable

Smith, Dominic

Published in:
New Formations

DOI:
10.3898/NEWF:98.05.2019

Publication date:
2020

Document Version
Peer reviewed version

[This is the penultimate version of an article published in *New Formations* 98 (2020) (

DOI: 10.3898/NEWF:98.05.2019)

It may contain differences from the published version. Please

do not quote this material without my written permission. Bracketed numbers

indicate pagination in the published version.]

Making Automation Explicable: A Challenge for Philosophy of Technology

Abstract: This essay argues for an expanded conception of automation's 'explicability.' When it comes to topics as topical and shot through with multifarious anxieties as automation, it is, I argue, insufficient to rely on a conception of explicability as 'explanation' or 'simplification.' Instead, automation is the kind of topic that is challenging us to develop a more dynamic conception of explicability as *explication*. By this, I mean that automation is challenging us to develop epistemic strategies that are better capable of *implicating* people and their anxieties about automation in the topic, and, counterintuitively, of *complicating* how the topic is interfaced with. The essay comprises an introduction followed by four main parts. While the introduction provides general context, each of the four subsequent parts seeks to demonstrate how diverse epistemic strategies might have a role to play in developing the process just described. Together, the parts are intended to build a cumulative case. This does not mean that the strategies they discuss are intended to be definitive, however – other strategies for making automation explicable may be possible and more desirable.

Part one *historicises* automation as a concept. It does this through a focus on a famous passage from Descartes' Second Meditation, where he asks the reader to imagine automata glimpsed through a window. The aim here is to rehearse the presuppositions of a familiar 'modernist' epistemological model, and to outline how a contemporary understanding of automation as a wicked socio-economic problem challenges it. Parts two and three are then framed through concepts emerging from recent psychology: 'automation bias' and 'automation complacency.' The aim here is to consider recent developments in philosophy of technology in terms of these concepts, and to dramatically explicate key presuppositions at stake in the form of *reasoning by analogy* implied. While part two explicates an analogy between automation bias in philosophical engagements with technologies that involve a 'transcendental' tendency to reify automation, part three explicates an analogy between automation complacency and an opposed 'empirical turn' tendency in philosophy of technology to privilege nuanced description of case studies. Part four then conclude by arguing that anxieties concerning automation might usefully be redirected towards a different sense of the scope and purpose of philosophy of technology today: not as a movement to be 'turned' in one direction at the expense of others ('empirical' vs 'transcendental', for instance) but as a multidimensional 'problem space' to be explicated in many different directions at once. Through reference [p 69] to Kierkegaard and Simondon, I try to show how different approaches to *exemplification, indirection* and *indeterminacy* can be consistent with this, and with the approach to explicability recommended above.

Keywords: *automation; automation bias; automation complacency; explicability; margin of indeterminacy; Kierkegaard; Simondon.*

Philosophically engaging a topic like automation today can bring anxieties that abstractions have been reified, or of becoming a 'Cassandra', 'essentialist', or 'technological determinist.' Turning to empirical complexities surrounding the topic can, conversely, bring anxieties that more global issues

are being ignored in favour of localised case studies, and concerns of over-specialism, positivism, or of being out of date in a fast-moving field. Whichever way, anxieties concerning 'topicality' emerge: is one seizing on a zeitgeist issue as a vehicle for fashionable dystopianism or utopianism, or, by being narrowly 'topical', is one overlooking bigger issues?

This essay explores such anxieties, with a view to how contemporary philosophy of technology might help address them. The key reason for doing so can be highlighted through an example drawn from the related field of AI research. In their recent paper, 'AI4People', Floridi et al. argue for five principles 'that should undergird [the] development' of 'AI for society.'[1] The first four are taken from bioethics, because, '… of all areas of applied ethics, bioethics is the one that most closely resembles digital ethics in dealing ecologically with new forms of agents, patients, and environments.'[2]Additionally, however, a fifth principle is argued for: 'explicability' ('AI4People', p696). By this, Floridi et al. mean 'the need to *understand* and *hold to account* the decision-making processes of AI', 'on grounds that are readily understandable to the proverbial person 'on the street.'' ('AI4People', p701).

Explicability emerges here as something additive. While apparently innocuous, this is vexed. This is because key epistemological issues are already at stake in Floridi et al.'s initial choice and justification of bioethics as a comparable ethical field.[3] In principle, these issues should have been explicated before the other four principles were imported over from bioethics. In fact, however, they remain implicit throughout Floridi et al.'s account.[4]

This essay seeks to demonstrate that explicability is a much more central concern than this. By reference to historical and contemporary philosophy, I want to show how explicability in the case of a topic like automation is not reducible to *explanation* or *simplification* for the 'proverbial 'person 'on the street.'' Instead, our concept of explicability should also aim at more thoroughly *implicating* people in automation as a topic, through a process of *explicating, complicating and critiquing* the interfaces through which diverse aspects of the topic are articulated and encountered.[5]

This process can begin with the concept of the 'person 'on the street.'' To [p 70] draw this out, part one foregrounds a famous passage where Descartes pictures people on a seventeenth century street. The aim here is to briefly rehearse a familiar 'modernist' epistemological model, and to show how automation challenges it.[6] Parts two and three are then framed through concepts emerging from recent

---

[1] Luciano Floridi, Josh Cowls, Monica Beltrametti, et al. 'AI4People', *Minds & Machines* (2018) 28: 689 – 707. (https://doi.org/10.1007/s11023-018-9482-5) (Hereafter, 'AI4People').

[2] The four principles are: 'beneficence', 'non-maleficence', 'autonomy', 'justice.'

[3] For instance: reasoning by analogy between ethical fields, and concepts of 'agency', 'patiency' and 'environment.'

[4] A more egregious example occurs in Iyad Rahwan et al. 'Machine Behaviour', *Nature*, 568 (2019), 477–486 (https://www.nature.com/articles/s41586-019-1138-y). Whereas Floridi et al. foreground an analogy with bioethics that remains underexplained, Rahwan et al. inflate an analogy into an equivalence (between an argued for 'interdisciplinary study of machine behaviour' and Nikolaas Tindbergen's approach to ethology).

[5] 'Ex*pli*cability', 'sim*pli*fication', 'im*pli*cation', 'ex*pli*cation' and 'com*pli*cation' each involve a sense of the '*pli*' as 'folding'. The claim here is that automation is the kind of challenge requiring us to develop a more baroque epistemology that does not merely privilege 'clarity and distinctness', in order to unfold its im*pli*cations (see Gilles Deleuze, *Le Pli: Leibniz et le baroque*, Paris, Seuil, 1988). By 'interface' here, I mean the media 'through which reality is experienced and interpreted', and have in mind a broad sense of media (for instance: epistemological models, dramatic *mise en scenes*, concepts, examples, analogies and images) (See Luciano Floridi et al., 'The Onlife Manifesto', in Luciano Floridi et al., *The Onlife Manifesto: Being Human in a Hyperconnected Era*, Springer: London, 2015, 7).

[6] See Bruno Latour, *We Have Never Been Modern*, trans. C. Porter, Cambridge, MA, Harvard University Press 1993.

psychology: 'automation bias' and 'automation complacency.' The aim here is to consider recent developments in philosophy of technology in terms of these concepts, and to dramatically explicate key presuppositions at stake in the comparison. While part two explicates an analogy between *automation bias* in philosophical engagements with technologies that involve a 'transcendental' tendency to reify automation, part three explicates an analogy between *automation complacency* and an opposed 'empirical turn' tendency in philosophy of technology to privilege nuanced description of case studies. Part four then concludes by arguing that anxieties concerning automation might usefully be redirected towards a different sense of the scope and purpose of philosophy of technology today: not as a movement to be 'turned' in one direction at the expense of others ('empirical' vs 'transcendental', for instance) but as a multidimensional 'problem space' to be *explicated* in many different directions at once.[7] Through reference to Kierkegaard and Simondon, I try to show how different approaches to *exemplification*, *indirection* and *indeterminacy* can be consistent with this, and with the approach to explicability recommended above.

Cartesian *Sangfroid*

Halfway through the Second Meditation, Descartes gives a famous glimpse of 'automata' as a philosophical problem:

> I marvel at how prone my mind is to errors … [W]ere I perchance to look out my window and observe men crossing the square, I would ordinarily say I see the men themselves. … But what do I see aside from hats and clothes, which could conceal automata. Yet I judge them to be men. Thus what I thought I had seen with my eyes, I actually grasped solely with the faculty of judgement, which is in my mind.[8]

As part of his work in anatomy, Descartes developed a theory of reflex action that is consistent with an archaic sense of 'automation', as 'automatism.'[9] Descartes was comfortable with the notion that embodied human beings are 'automated' in this sense: if I place my foot in a fire, I will remove it through automated reflex, not reflective judgement. This view of the human body, he thought, was consistent with his dualistic metaphysics of the human *being*. For Descartes, the human being could not be fully accounted for in terms of a body: it was a 'comingling' of an automated body and a free and rational mind.

This 'substance dualism' is widely discredited today. On the account developed in this essay, this is not merely due to a familiar story about [p 71] how developments in scientific naturalism and secularism have eroded dualism's metaphysical plausibility. It also has to do with a shift in the kind of philosophical problems that 'automata', 'automation' and 'automatism' are posing. For Descartes in the *Meditations*, these were topic to be glimpsed through an epistemological model having to do with how a subject recognises other beings with minds. Today, this model has been superseded by a sense of automation as a 'wicked' socio-economic problem. Simply: what happens when 'automata' can outperform and replace human beings across all kinds of socio-economic contexts, irrespective of whether they have 'minds' or not? For Descartes, this problem was barely glimpsed. At the beginning of the twenty-first century, however, it has become anxiously topical.

---

[7] By 'empirical', I mean 'concerned with the actual or possible objects of experience'. By 'transcendental', I mean 'concerned with conditions for the possibility of experience'. My claim is that we should be open to emergences in both of these directions, at once.

[8] René Descartes, *Meditations, Objections, and Replies*, trans. R. Ariew and D. Cress, Indianapolis, Hackett, 2006, 32.

[9] 'Automation, n.' *OED Online* (www.oed.com/view/Entry/13468).

Having alluded to his problem concerning automata, Descartes quickly shuts it down through a judgement offered with apparent *sangfroid*: 'I judge them to be men.' But there is a paradox here: Descartes felt justified to state *that* he judged this way without stating *why* and *how*, and so he responds to the problem of automata with a self-assurance that is itself 'automatic' by his own epistemological model, in the sense of unreflective.

Today, we are no longer in the position of disinterested Cartesian subjects musing about possible automata in the square, and nor, it seems, are we in a position to respond to problems posed by automation with unreflective *sangfroid*. Instead, we are beings whose everyday socio-economic reality is *implicated* in complex technological networks that connect up more or less acutely to issues concerning automation. To the extent that we grasp this, moreover, it is less likely to occur transparently, as if through windows onto a square. Instead, it is more likely to occur partially to even the most expert among us, through interfaces that render any sense of 'the whole' more or less opaque, unstable, and perhaps counterproductive.[10]

Irrespective of how well-founded Descartes's apparent *sangfroid* was, then, we appear to have lost it in the face of a cluster of emergent socio-economic problems concerning automation. This means, first, that we are less able than he was to displace issues concerning why and how 'automata' are problematic, on pain of having them return as overwhelming anxieties, and, second, that we are *implicated* and *complicit* in the topic in ways that challenge us to develop new forms of explication for it.

Automation Bias

If our sense of the type of problem posed by automation has shifted from a Cartesian epistemological model centred on the reflective subject as independent, so too should our sense of philosophical *mise en scene.* In this part and the following one, I will take up two concepts from recent psychology in an attempt to explicate some consequences of this for philosophy of technology today: 'automation bias' and 'automation complacency.'

Automation bias is 'the tendency to over-rely on automation' when using [p 72] automated support systems, as manifested both 'in errors of commission (following incorrect advice) and omission (failing to act because of not being prompted to do so).'[11] As Carr elaborates:

> [Automation bias] creeps in when people give undue weight to the information coming through their monitors. Even when the information is wrong or misleading, they believe it. Their trust in the software becomes so strong that they ignore or discount other sources of information, including their own senses.[12]

Consider Carr's reference to 'trust in the software.' In this part, I want to explore the sense in which this might be analogous to a classical or 'transcendental' tendency in philosophy of technology: to reduce given technological phenomena to an abstract account of their putative conditions of

[10] See Alexander Galloway, 'Are Some Things Unrepresentable?', *Theory, Culture & Society*, 28, 7-8 (2011), 85-102, and Wendy Hui Kyong Chun, *Programmed Visions: Software and Memory*, Cambridge MA, MIT Press, 2011, p1-13.
[11] Kate Goddard, Abdul Roudsari, and Jeremy C Wyatt, 'Automation Bias: A Systematic Review of Frequency, Effect Mediators, and Mitigators', *Journal of the American Medical Informatics Association : JAMIA*, 19, 1 (2012): 121-127. (Hereafter 'Automation Bias').
[12] Nicholas Carr, *The Glass Cage: Where Automation is Taking Us*, London, The Bodley Head, 2015, p67. (Hereafter *The Glass Cage*).

possibility. The canonical example of this is Heidegger's approach, which considers disparate technological phenomena as instances of the 'way of revealing' called 'Enframing' ('*Gestell*.')[13] Viewed in terms of this analogy, Carr's emphasis on discounting our 'own senses' is also important: since an 'empirical turn' in philosophy of technology in the 1990s, the accusation levelled against 'classical' approaches is that, by indulging a theoretical bias towards transcendental abstraction, they ignore the need for a more empirical approach to the complexities of specific technological phenomena.[14]

Can we point to more contemporary instances of this tendency, and how does it relate to automation? Consider the following passage from Srnicek and Williams's *Inventing the Future*:

> The most recent wave of automation is poised to change [the] distribution of the labour market drastically, as it comes to encompass every aspect of the economy: data collection (radio-frequency identification, big data); new kinds of production (the flexible production of robots, additive manufacturing, automated fast food); services (AI customer assistance, care for the elderly); decision-making (computational models, software agents); financial allocation (algorithmic trading); and especially distribution (the logistics revolution, self-driving, drone container ships and automated warehouses). In every single function of the economy - from production to distribution to management to retail - we see large-scale tendencies towards automation. This … is predicated upon algorithmic enhancements (particularly in machine learning and deep learning), rapid developments in robotics and exponential growth in computing power (the source of big data) that are coalescing into a 'second machine age' that is transforming the range of tasks that machines can fulfil.[15]

Consider the last sentence here, where possibility rather than necessity is invoked (a 'can' rather than 'must'). This sense of possibility runs right through [p 73] *Inventing the Future's* argument: that the advent of a 'second machine age' makes it possible for us to transcend the category of labour as traditionally configured, and that we should be politically committed to this historically contingent possibility. On Srnicek and Williams's account, it is precisely by recognising what is at stake in this possibility that contemporary human beings will be able to invent a tolerable future. What is noteworthy, however, is how this jars with everything else in the passage which, modally speaking, gives the impression that we are firmly in the domain of necessity.

Consider Srnicek and Williams's list of 'aspects of the economy.' The least that can be said about it is that it is compressed and cursory. More problematically, it tends towards a kind of rhetorical steamroller effect. Of itself, this needn't be problematic: it might work as a kind of 'shock to thought' before a more applied account of similarities and differences between the various aspects of the economy Srnicek and Williams describe, along with an account of how these can be interrelated in terms of the problem of automation. What is problematic about Srnicek and Williams' approach, however, is that this is precisely what is missing.[16]

---

[13] Martin Heidegger, *The Question Concerning Technology and Other Essays*, trans. W. Lovitt, New York, Harper & Row, 1977.
[14] See Hans Achterhuis (ed), *American Philosophy of Technology: The Empirical Turn*, trans. R. P. Crease. Bloomington, Indiana University Press, 2001.
[15] Nick Srnicek and Alex Williams, *Inventing the Future: Postcapitalism and a World Without Work*, London, Verso, 2015, pp110-111.
[16] See Alexander Galloway, 'Brometheanism', *Culture and Communication*, 16 June 2017, http://cultureandcommunication.org/galloway/brometheanism.

The relation between the domains listed by Srnicek and Williams is, quite simply, much more complex, contingent and subject to temporal 'lag' and disjunction than they make it appear. Where a recognition and account of this is missing, however, what stands to fill the gap is a reified sense of 'automation' as transcendental force acting homogeneously, unilaterally, and synchronously across 'every single function of the economy.' This, I suggest, an updated version of the transcendental tendency to reify 'Technology' (a kind of 'transcendentalism 2.0'), and is problematic not merely for Srnicek and Williams, but for many contemporary approaches that attempt, like them, to theorise automation in more or less speculative terms: following a rhetorical shock to thought involving multiple examples, further analysis is typically missing, and this opens a space for the reification of 'automation' into an englobing force capable of 'filling the gaps.'[17]

This tendency can, I think, be read as analogous to a form of 'automation bias' implicit in philosophical engagements with technology that trust too much to the 'software' of speculation and theory *themselves*. It is problematic both in principle and in fact: in principle because it tends to hypostasise and divorce theoretical abstractions from emergent empirical complexities; in fact because it gives the impression that competency in theory is sufficient to replace engaged praxis within situated problems.

The main problem here, however, may be more one of fact than principle. Reconsider Srnicek and Williams's list: what is problematic is not so much the issue of principle (grouping diverse aspects of the economy in terms of automation), but instead the fact that Srnicek and Williams do not go on to explore aspects of the economy so listed. What is problematic about this in general is that it leaves the way open for 'transcendentalism 2.0.' What is problematic in terms of Srnicek and Williams's approach specifically is that [p 74] it obscures an opportunity to make good on the politically activist principle of their approach.

Making these points is not to accuse Srnicek and Williams of being insufficiently 'expert', and need not undermine their principle of a politics committed to understanding automation. It is rather to note a far richer range of political possibilities for making automation *explicable* that stand to emerge when we resist the tendency to *explain* the topic in terms of the interface of 'transcendentalism 2.0' and instead undertake to *explicate* the complex relations between the 'aspects of the economy' that Srnicek and Williams bunch together. This, by its nature, will involve a more 'bottom up' than 'top down' approach, but need not lead back to the forms of 'localism', 'folk politics' and political 'common sense' that are provocatively critiqued in *Inventing the Future* (pp5-23). On the contrary, it can be part of a scalable project aimed at complicating the types of experiences and expertise that are given voice. What, for instance, are the instructive *differences* between automation as implemented in customer assistance and care for the elderly? What *isomorphisms* stand to emerge across different aspects of the economy when we resist a tendency to see relations in terms of reductive patterns of unidirectional causality? What *different impacts* do factors such as neural networking, crowdsourcing and big data have for different aspects of the economy?[18]

---

[17] See, for instance: Antonio Negri, 'Reflections on the "Manifesto for an Accelerationist Politics"', *E-flux*, 53 (2014); Bernard Stiegler, *La Société automatique: 1. l'avenir du travail*, Paris, Fayard, 2015; Franco Berardi, *Futurability: The Age of Impotence and the Horizon of Possibility*, London, Verso, 2019; Matteo Pasquellini, 'Capital Thinks Too: The Idea of the Common in the Age of Machine Intelligence', *Open! Platform for Art, Culture & the Public Domain*, 11 December 2015, http://www.onlineopen.org/capital-thinks-too.

[18] Consider, for instance, the different impacts of automated systems upon customer service and care of the elderly, with respect to loneliness and social isolation. In a survey of eight different types of automated systems, across fifteen studies from 2002 to 2014, including pet robots, conversational agents, and 'personal reminder information and social management systems', Khosravi et al. report a 'decrease in social isolation and loneliness

Automation Complacency

I have just argued that a tendency towards 'transcendentalism' may be more problematic in fact than in principle. This presupposes that speculative developments in 'theory' can be valuable for making a topic like automation explicable, as a matter of principle. The aim for this part is to contextualise this point in terms of a tendency emerging in philosophy of technology since an 'empirical turn' in the late 1990s. While some of the approaches I will discuss aim to go beyond the empirical turn, they prioritise specific engagements with technologies in contexts of design, implementation and use, and differ from the approach just outlined in that they take tendencies towards 'transcendentalism' to be problematic *in principle*. In this part, I will suggest that these approaches involve something analogous to a form of 'automation complacency' that overlooks what might be valuable about a developed and nuanced sense of the transcendental.

As defined by Goddard *et al*, automation complacency involves 'insufficient monitoring of automation output' ('Automation Bias', p121). As Carr puts it:

> [Automation complacency] takes hold when a computer lulls us into a false sense of security. We become so confident that the machine will work flawlessly, handling any challenge that may arise, that we allow our attention to drift. We disengage from our work, or at least from the part [p 75] of it that the software is handling**,** and as a result may miss signals that something is amiss (*The Glass Cage*, p67).

There are obvious overlaps with 'automation bias' here. A key difference, however, concerns the place of conscious attention in 'human factors' monitoring automated processes. In cases of automation bias, there is more scope for *conscious inattentiveness* to discount conflicting inputs and outputs; in cases of 'automation complacency', the emphasis is on forms of inattention that have become *habitual and unconscious*. In terms of the analogy I will develop in this part, this difference can be compared to that between philosophical approaches that are consciously theory-driven and speculative, and approaches that, in aiming to be as empirical as possible, tend to overlook how mediated by theory they may in fact be.

Consider Carr's remark that automation complacency can lead us to 'miss signals that something is amiss.' In this part, I will suggest that something comparable may be 'amiss' in the way certain empirical turn approaches have restricted their focus. As a way in, consider the following from Verbeek:

> [T]raditional philosophy of technology approached its subject matter from a *transcendental* direction. Transcendental philosophy, which achieved its zenith in the work of Immanuel Kant, takes as its point of departure the analysis of conditions of possibility … This approach has produced many relevant insights and to a large extent has shaped the understanding of

among seniors', especially in robotics use (Pouria Khosravi, Azadeh Rezvani, and Anna Wiewiora, 'The Impact of Technology on Older Adults' Social Isolation', *Computers in Human Behaviour* 63, C, (2016): 599). In contrast, automated checkouts are reported to have marked detrimental effects upon opportunities for work among women of all ages in caregiving roles, exacerbating problems of social isolation (Robert Booth, 'Shop Closures and Self-Checkouts Cost Tens of Thousands of Women's Jobs', *Guardian*, 17 September 2019, https://www.theguardian.com/money/2019/sep/12/online-shopping-forcing-women-out-of-work-study). Srnicek in particular has been active in pursuing this kind of more granular analysis through online fora (see, for instance: 'Autonomy' 17 September 2019, https://autonomy.work/) What is required is not merely a use of different fora to *explain* automation in terms of 'transcendentalism 2.0', however, but rather use of different fora to *explicate* the topic, in terms of how we are differently implicated in it (as caregivers or users of self-service checkouts, for instance).

> technology and its role in contemporary culture. But our picture of technology is distorted if technology is approached *exclusively* in terms of its conditions … For then we are speaking about technology's conditions of possibility as if we were speaking about concrete technologies themselves, and the transcendental perspective becomes absolutised into *transcendentalism*. This is precisely what happens in classical philosophy of technology … The philosophy of technology needs to resist this 'Orphic temptation' of looking backward. It must be confident that it will be able to get a full view of technology once it has left the realm of the transcendental and re-enters the world of concrete materiality.[19]

Verbeek starts with an insightful summary of what the transcendental approach in philosophy involves. But note a slippage: having described the transcendental in methodological terms, as an 'approach' and 'perspective', he closes by describing it in metaphysical terms, as a 'realm.' This, I suggest, is symptomatic of what is amiss in how empirical turn approaches treat the transcendental: by reifying it into an otherworldly realm out of touch with 'concrete materiality.'

This tendency is problematic in general because it blocks us from looking to the history of philosophy to develop comparative approaches that engage a sense of the transcendental, not as an otherworldly realm, [p 76] but as a methodological problem concerning the scope and relevance of conditions in philosophical inquiry. The tendency exemplified by Verbeek's remarks is also problematic in terms of automation specifically, however. This is because it seems to imply a relatively fixed sense of what constitutes a 'technology', at precisely the moment when issues surrounding automation are challenging this.

But couldn't it be objected that I have taken Verbeek's words in isolation here, or that the 'empirical turn' is a dated matter?

Since the empirical turn, there have been many calls for further 'turns' in philosophy of technology.[20] Instead of being viewed as a dated event, then, the empirical turn can be viewed as a precedent that says something about how a picture of method as 'turning' has held across philosophy of technology. To see what might be amiss with this, consider the following from Peter Kroes, a proponent of the 'engineering turn':

> All the various forms of technology … may hang together through … family resemblance, without there being a common core element. Modern technology is a historically grown, highly complex and diverse phenomenon, a fact that should not be ignored by philosophers of technology. This is only possible through a shift from a global to a more local level of analysis. The richness of technology will become visible only by looking at modern technology through a magnifying glass.[21]

Kroes starts with an interesting diagnosis of the problem of technological complexity. But note how exclusive his prescription is. Quite simply, why is it 'only possible' to understand modern technology by shifting from the global to the local?

---

[19] Peter-Paul Verbeek, *What Things Do: Philosophical Reflections on Technology, Agency and Design*, Philadelphia, Penn State Press, 2005, p7.
[20] See Dominic Smith, *Exceptional Technologies: A Continental Philosophy of Technology*, London, Bloomsbury, 2018, pp107–127. (Hereafter *Exceptional Technologies*).
[21] Peter Kroes, 'Engineering Design and the Empirical Turn in the Philosophy of Technology', in Kroes, P. and A. Meijers (eds), *The Empirical Turn in the Philosophy of Technology*, Bingley, Emerald Group, 2000, p28.

If it was logically impossible to attempt a global level of analysis, there would be less reason for Kroes to warn against it. The force of his remarks is therefore normative, and bound to a particular picture of method as 'turning.' According to this picture, every 'turn towards' must involve a 'turn away' from something else. The 'empirical turn' in philosophy of technology was, as suggested above, a precedent instance of this logic insofar as it prescribed a turn towards the empirical and a turn away from the transcendental; following in the wake of the empirical turn, Kroes's prescription of the local over the global also exemplifies this logic.

How might it be possible to go further in *both* directions described by Kroes, at once, towards a responsive interrogation of concepts of the global and the local alike, and of their interrelations, in favour of allowing cases and topics to show up that are occluded by our received sense of these categories? According to the picture of turning, such an endeavour is contradictory. But what if this picture was replaced (or, better, *emplaced*) in terms of different pictures of method?[22] Then the appearance of contradiction might disappear in favour of changed coordinates and new directions for explication.[23] [p 77]

Recognising this possibility is not simply to play out one more dialectical permutation in a language game of 'turning.' Nor is it a 'transcendental' turn. It is to suggest that philosophy of technology cannot remain in complacent recourse to a picture of 'turning' in the face of contemporary challenges posed by topics like automation. This is because such topics show up as paradoxical exceptions to the received coordinates of this picture: as at once 'global' and 'local', metaphysical/epistemic and socio-economic, theoretical and existential, 'transcendental' and 'empirical.' Thereby, they challenge us to develop new forms of explication.

Implicated/Explicated

I highlighted some ways in which automation's 'topicality' has become problematic at the beginning of this essay. The challenge this poses for philosophy of technology specifically is not to polarise in opposed 'transcendental' and 'empirical' directions. Today, automation is at once a highly abstract, speculative and englobing topic, yet empirically nuanced, discrete and localised. Rather than being reducible to transcendental biases or empirical complacencies, then, it is the kind of topic that challenges us to question these biases and complacencies *alike*, in favour of a sense of philosophy of technology as a complex problem space, with potential to be explicated in many different directions at once.

My aim here is to try to exemplify how this space might be explored. Rather than proceeding in a 'transcendental' or 'empirical' direction, I will proceed through a method of *indirection*. This may initially appear tangential, and will likely appear too promissory or partial in its sources. By the end, however, I hope to have dispelled what is negative about these impressions, in favour of a sense of a space having 'opened up.'

In *Either/O*r, Kierkegaard writes:

[22] For instance: as 'mapping', 'topography' or 'topology.' See *Exceptional Technologies*, p27–33.
[23] It is perfectly possible to go in many different directions at once according to different pictures of method, by changing the relative coordinates of the space surveyed. One does it when zooming in and out on a digital map interface.

> My reflection on life altogether lacks meaning. I take it some evil spirit has put a pair of spectacles on my nose, one glass of which magnifies to an enormous degree, while the other reduces to the same degree.[24]

So far in this essay, we have encountered striking images from thinkers both historical and contemporary: Descartes's automata, Srnicek and Williams's 'wave' of automation, Verbeek's Orpheus, Kroes's 'magnifying glass.' I now want to suggest that, short, refractory, and remote from contemporary concerns on automation though it seems, Kierkegaard's image explicates important aspects of this topic that the others overlook. This is not because it describes the reality of our situation 'better' in a metaphysically absolute sense. It is because it is the kind of exceptionally aporetic image that challenges us to reconsider what we take the reality of our situation to be, and to explicate it differently, in terms of new ways of interfacing with it. [p 78]

Descartes's automata were meant to be glimpsed through a window: a view onto a square implying distance, 'overview', a private/public separation, and an assumed 'transparency.' In contrast, Kierkegaard's lenses highlight the extent to which different media and interfaces can appear transparent in isolation, but have distorting effects when functioning together. If, as Kierkegaard's image suggests, this can be the case with two analogue visual media of the same kind, what happens when we are situated in terms of a proliferating array of networked media and interfaces, each appealing to different sensory modalities and blends thereof, and each connected more or less acutely to the stakes of automation as a multifarious socio-economic topic?

Srnicek and Williams's wave gave focus to automation as a 'clear and present danger' set to overwhelm. In contrast, Kierkegaard's image implies a lived existential anxiety of indefinite proportions: less an imminent inundation from everywhere at once, more a coastal erosion or series of sinkholes yet to surface. In contrast to Srnicek and Williams's sophisticated litany of 'aspects of the economy', Kierkegaard's image is as precise and accessible as it is bizarre.

Verbeek's Orpheus was tempted back towards a 'nether' world by transcendental spirits. Kierkegaard's 'evil spirit' suggests he is not so naïve. Rather than being duped by powers beyond his control, Kierkegaard is, like Descartes before him, personifying a doubt (but in a different way, that is more lived and existential, less theoretical). Consider, in this respect, Kierkegaard's 'I take it': this implies self-conscious awareness of the tropes of language; it is not a credulous fallacy of reification, but an ironic *prosopopoeia*.

Kroes's magnifying glass implies the greatest superficial proximity to Kierkegaard's image through its ocular emphasis. They part company, however, in the emphasis that Kroes places on a common sense of vision, in contrast to Kierkegaard's emphasis on paradox. Viewed in terms of Kroes's image, it is as if Kierkegaard were describing the effects of looking through a magnifying glass and a telescope at once. Viewed in terms of Kierkegaard's image, it is as if Kroes were describing the effects of someone with binocular vision choosing to examine the world through one eye.

Automation today is at once a topic implying multiple analogue and digital media across different sensory modalities. It is a topic that is at once linked to contemporary existential anxieties of indefinite extent, and something encountered in everyday experience in precise yet bizarre ways. And it is at once a topic that the historically evolved tropes of our natural languages can push us towards reifying and personifying, and a kind of all-pervasive atmosphere involving no overseeing 'super-intelligence.' Excepting Descartes's image, Kierkegaard's is the most historically remote of those

[24] Søren Kierkegaard, *Either/Or*, trans. A. Hannay, Harmondsworth, Penguin, 1992, p46.

considered in this essay. As the issues just touched on suggest, however, it may nevertheless be the kind of aporetic interface we need today. This is because, in a climate where automation has become anxiously 'topical', such an image has the capacity to act as a focal point for explicating [p 79] untimely aspects of the topic, not in spite of its apparent remoteness, but because of it.

Suppose we consider Kierkegaard's spectacles as a kind of interface between the lived subjective effects of automation and broader socio-economic complexities implied by this topic. Quite apart from whether such an interface is possible, a sign it is desired today is the preponderance of contemporary theorising devoted to the topic. What strategies does Kierkegaard's image suggest for mediating this desire?

One strategy would be to engage automation in precise practical ways in some situations, and as a speculative challenge in others. This could be viewed like viewing the world through a magnifying glass, then a telescope, alternately. Framed in terms of Kierkegaard's image, however, such a strategy might quickly show up as maladaptive: like someone with binocular vision trying to navigate the world with one eye closed at a time.

Another strategy would be to take the spectacles off. This could be viewed as analogous to choosing to disengage from the topic. Even if such a choice might appear necessary or desirable as a form of intermittent therapy, however, it might begin to look like evasion from the topic's 'topicalities': like a choice that leaves us myopic when we feel clear-sighted.

What if automation is the kind of topic that is challenging us to adopt the least likely strategy implied in Kierkegaard's image? That is, something analogous to navigating the world as subjects of binocular vision, looking through apparently contradictory lenses, with *both* eyes open at once?

This would be the most puzzling and uncomfortable strategy, both in terms of someone trying to live through it, and in terms of its analogical implications for Kierkegaard's philosophy. As a thinker of the existential 'either/or', Kierkegaard emphasised the necessity of choice and an anguished 'leap to faith' in the face of mutually exclusive options. Wouldn't the strategy just suggested be entirely at odds with this?

But note how Kierkegaard frames his image: he is in a situation where his 'reflection on life lacks all meaning.' What he is describing is not a choice between two mutually exclusive decision paths (either the right lens or left), but the disorienting effects of a breakdown of the received framework in terms of which choice, *per se*, seemed possible. What Kierkegaard's image implies, in this sense, is not a localised challenge to meaning, but a challenge to an entire picture in terms of which meaning and choice seemed possible.

The suggestion I want to make in closing this essay is that automation may be the kind of topic that is posing this kind of challenge today, in all sorts of ways (whether more or less technical/theoretical or 'everyday'). This is not to say that automation is the only or most fundamental topic posing this kind of challenge, or that it amounts to a new form of '*Gestell*' or 'transcendentalism 2.0', destined to monotonously englobe our sense of literally everything else. On the contrary, a key challenge facing philosophy of technology today is that there may be indefinitely many other such topics that implicate

us in [p 80] this way, each requiring the development of different ways of rendering them explicable.[25]

One potentially important strategy for beginning to explicate these topics, I suggest, is to seek out the kind of aporias that show up as tangential or exceptional to received forms of understanding on them. Put simply: what kind of aporias can we find to expose and frustrate the limits of forms of understanding that may themselves have become too 'automated', in favour of a sense of understanding as a shared and renewable project of explication, capable of being undertaken in common, and involving a multiplicity of different stakeholders and factors?

Kierkegaard's image is, I suggest, one such aporia in the face of our sense of automation's 'topicality.' But isn't it possible that, rather than providing a different or 'untimely' way in to the topic, it won't speak to us at all, or that it simply deflects attention away from the 'real' and 'technical' problems at hand?

Assume that something analogous to the 'both eyes' strategy implied in Kierkegaard's image could be adopted. This would be like arriving at a technique for rendering us hyper-aware of the implications of automation in terms of micro and macro levels that at once go both beneath and above the thresholds of normal human sense perception. Wouldn't this be unliveable? At the least, it would seem to involve a state of constant 'high alert', as described by Carr in the conclusion of his reflections on automation complacency and bias:

> Both [automation] complacency and bias tend to become more severe as the quality and reliability of an automated system improve. Experiments show that when a system produces errors fairly frequently, we stay on high alert**.** We maintain awareness of our surroundings and carefully monitor information from a variety of sources. But when a system is more reliable, breaking down or making mistakes only occasionally, we get lazy. We start to assume the system is infallible (*The Glass Cage*, p71).

A pervasive received picture of what constitutes an automated system or process today dictates that it should tend towards ever increased efficiencies. A further received picture, perhaps pervasively felt, but only theorised on a more specialist level, holds that increased efficiencies may involve collateral effects of automation bias and complacency in 'human factors' attending an automated system. Further pictures hold that human factors may therefore have to be automated out of the picture entirely. These pictures are related in complex ways and not necessarily reducible to one another. In contrast to them all, however, Kierkegaard's image offers something else: the spectacles he describes are an interface for producing unreliability, fallibility and error; their 'product' is not efficient functioning, but a state of nausea and/or perpetual 'high alert.' [p 81]

What Kierkegaard's image offers when related to automation, in this sense, is something much more than a momentary sense of strangeness. It is rather a challenge to develop techniques for critically and creatively interrogating and explicating our received pictures of automated systems and processes in non-reductive ways. Assuming such a challenge could be taken up, however, wouldn't the implied state of 'high alert' still be unliveable?

As discussed, there are at least three strategies implied in Kierkegaard's image: looking through one eye at a time, taking the spectacles off, and both eyes at once. Of these, the 'both eyes' strategy might

[25] For instance: the Internet of Things, biotechnologies, nanotechnologies, the semantic web, cryptocurrencies, the Anthropocene.

be the most philosophically challenging. This does not exclude the possibility of adopting something analogous to the others, however. A first response to the problem of 'high alert', then, would be to note that while we could conceivably be in such a state all the time, we wouldn't necessarily have to be. Instead, we might render this state more liveable by adopting different strategies in response to different aspects of the problem.

A second response involves emphasising a possibility, less of 'adoption', than 'adaptation.' Put simply, a state of 'high alert' might only appear this way relative to habitually acquired forms of life. Given sufficient exposure to new conditions, we might conceivably adapt to them. Viewed in terms of Kierkegaard's image, this would be like reaching the point where we can see through both lenses at once, without a sense of nausea.

This response is suggestive, but vexed. To draw this out, consider the following from Gilbert Simondon, on his concept of the 'margin of indeterminacy':

> Automatism is a rather low degree of technical perfection … [T]o render a machine automatic, it is necessary to sacrifice many possibilities of functioning, many possible usages. Automatism, and its use under the industrial form of organisation that we call *automation*, possesses more economic and social significance than technical significance. The true perfection of machines, that which raises the degree of technicity, corresponds not to a growth of automatism, but, on the contrary, to the fact that the functioning of a machine conceals [*receler*] a certain margin of indeterminacy. It is this margin that permits the machine to be sensible to external information. It is by this sensibility of machines to information that a technical ensemble can be realised.[26]

By 'margin of indeterminacy', Simondon means the variable cluster of possibilities and potentials surrounding any given technological artefact or process. For instance: all the possible glitches it will encounter; all the unforeseen consequences its functioning will imply for the elements shaping it and the broader 'ensembles' it forms a part of; all the repercussions it is capable of having as an object of imagination or ideology; all the applications that are *impossible* for it. For Simondon, such a margin of indeterminacy is a [p 82] necessary structural feature. As part of the development of any technology, this margin is explicated and adjusted. It is inconceivable for Simondon, however, that such a margin should ever be exhausted. On the contrary, any conceivable technological artefact or process will, he holds, possess a shifting margin of indeterminacy specific to it. For Simondon, the margin of indeterminacy is, in this sense, a virtual cluster surrounding any given technology, and a condition for the possibility of it being able to form part of any broader 'ensemble' at all.

How does this relate to Kierkegaard's image and the problem of 'high alert'? Assume that fears and anxieties about living in such a state might be reactions contingent on our habitually acquired forms of life. If these fears and anxieties could be overcome, what might then emerge, on Simondon's reading, are new forms of life that might be better attuned to the 'margins of indeterminacy' involved in the artefacts and processes that compose automation as an 'ensemble' topic. The danger, however, is that we could be deluded here: rather than helping us arrive at an understanding of the conditions of automation, this approach might simply amount to adjustment to a contemporary ideology that

[26] Gilbert Simondon, *Du mode d'existence des objets techniques*, Paris, Aubier, 2012, p12. (Hereafter *Du mode*).

emphasises ‘adaptability’ and ‘flexibility’, and that erodes the grounds on which principled resistance might stand.[27]

But what if a different form of ‘therapy’ is possible, not as ‘adaptation’ or ‘adjustment’, but as the will and capacity to make things *explicable* in a challenging sense? This indicates a third possible response to the problem of ‘high alert.’ Put simply: how can we begin to approach anxieties surrounding topics like automation today, not to repress or ignore them in favour of a dominant ideology, but to better explicate the ‘margins of indeterminacy’ that at once make them possible, and that they can serve to distort?

In the introduction to *Du Mode d’existence des objets techniques*, Simondon outlines a general hope that his book might effect a ‘*prise de conscience*’ with respect to ‘technical objects’:

> The *prise de conscience* of the modes of existence of technical objects should be effected by philosophical thought, which finds before it in this undertaking a task analogous to that which it played in the abolition of slavery and the affirmation of the worth of the human person (*Du mode*, p9).

Simondon’s strategy here is, I suggest, epistemologically vexed. This is because he appears to be seeking out *one* key *prise de conscience* by way of *one* key analogy, the moral implications of which may be just as likely to block understanding of his rich ontology as to enable it. Simondon’s analogy was no doubt calculated to appear provocative. Rather than advancing his position, however, it might equally well backfire and draw misunderstanding and moral opprobrium upon it.

What may be more desirable, then, is a ‘multiplication’, ‘pluralisation’, [p 83] or ‘intensification’ of Simondon’s approach: rather than proceeding through one key analogy that deliberately courts controversy, perhaps we should be aiming at multiple *prises de conscience* that stretch the limits of received patterns of analogical reasoning, and that problematise what we take ‘philosophical thought’ on technical objects to be. Rather than deliberately courting controversy, and rather than seeking to explain things from the ‘top down’, this approach would aim at a more fallibilistic and experimental form of explication, situated in the midst of things (and their complications). Thereby, it might have better potential to be responsive for different audiences, and attuned to different ‘technical objects.’ Building on Simondon’s ontology, the key point here is that new *prises de conscience* concerning the margins of indeterminacy involved in technical objects should always be possible, because there will always be such margins. What may be required to draw this out, however, is a more reflexive and nuanced epistemology of explicability, capable of being more inventive and open in the search for the kinds of focal points that might allow us to develop shared forms of understanding.

In this part, I have sought to show how a short image from Kierkegaard might act in this way. This image may have appeared tangential. When considered in terms of Simondon’s approach, however, it emerges differently: as an ‘interface’ or ‘technical object’ with its own rich ‘margin of indeterminacy.’ By situating this image in terms of automation, I have attempted to explore its margin of indeterminacy in new ways. The aim in doing so has not been to advance it as sufficient to enable understanding of the ‘whole’ of this complex topic; it has been to suggest that such images are both possible and desirable, as ways of explicating the topic differently.

[27] See Alexander R. Galloway, ‘The Poverty of Philosophy: Realism and Post-Fordism’, *Critical Inquiry*, 39, 2 (2013): 347-366; Catherine Malabou, *Que faire de notre cerveau?*, Paris, Bayard, 2011.

Today, automation is the kind of topic that can be viewed in terms of Cartesian robots, waves set to inundate, underworld gods, or trying to see things with microscopic exactitude. Additionally, however, it can be viewed in terms of processes implicating varying degrees of the surreal, the physically painful, the disorienting, and the apparently contradictory: as if trying to navigate the world looking through a magnifying glass and a telescope at once. Taken in isolation, none of these images is sufficient for understanding the whole of 'automation.' Together, however, they form a constellation capable of sustaining critical contrasts, new directions, and new ways of trying to make the complexities of this topic explicable, however indirect to its central concerns they may initially have seemed.

Explicability, as discussed at the beginning of this essay, implicates senses of 'explanation' and 'explication.' Explanation is the dominant sense here, but it is always 'to' or 'for' someone else, and implies a relatively fixed and determinate sense of the knowledge relation thereby (for instance: 'explainer' → 'simplified expertise' → 'person 'on the street.'') Explication, in contrast, implicates these terms in a more fluid and dramatic relation, by opening up the topic at stake (its '*explanandum'* or '*explicandum'*) as something shared and less determinate. That is: as something in which we are *complicit* with both [p 84] human and nonhuman others, to be explored in terms of interfaces that can be complicated and surprising.

If the 'proverbial person 'on the street' is to actually be recognised as someone capable of expressing their implication in automation, and their anxieties about it, the topic cannot be left to explanation alone. Instead, philosophical approaches to technology today are being challenged to explicate a more dynamic sense of their own scope, purpose, materials, and of how and for whom such topics are to be engaged.